Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Structural Biology 148 (2004) 194–204
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

XMIPP: a new generation of an open-source image processing package for electron microscopy

C.O.S. Sorzano[a,c,*], R. Marabini[a,b], J. Velázquez-Muriel[a], J.R. Bilbao-Castro[a], S.H.W. Scheres[a], J.M. Carazo[a,b], A. Pascual-Montano[a]

[a] *Unidad de Biocomputación, Centro Nacional de Biotecnología (CSIC), Campus Universidad Autónoma s/n, 28049 Cantoblanco, Madrid, Spain*
[b] *Escuela Politécnica Superior, Univ. Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain*
[c] *Dept. Ingeniería de Sistemas Electrónicos y de Telecomunicación, Univ. San Pablo-CEU, Campus Urb. Montepríncipe s/n, 28668 Boadilla del Monte, Madrid, Spain*

Received 30 March 2004; received in revised form 4 June 2004
Available online 14 August 2004

Abstract

X-windows based microscopy image processing package (Xmipp) is a specialized suit of image processing programs, primarily aimed at obtaining the 3D reconstruction of biological specimens from large sets of projection images acquired by transmission electron microscopy. This public-domain software package was introduced to the electron microscopy field eight years ago, and since then it has changed drastically. New methodologies for the analysis of single-particle projection images have been added to classification, contrast transfer function correction, angular assignment, 3D reconstruction, reconstruction of crystals, etc. In addition, the package has been extended with functionalities for 2D crystal and electron tomography data. Furthermore, its current implementation in C++, with a highly modular design of well-documented data structures and functions, offers a convenient environment for the development of novel algorithms. In this paper, we present a general overview of a new generation of Xmipp that has been re-engineered to maximize flexibility and modularity, potentially facilitating its integration in future standardization efforts in the field. Moreover, by focusing on those developments that distinguish Xmipp from other packages available, we illustrate its added value to the electron microscopy community.

Keywords: Electron microscopy; Single particles; Software packages; Image processing; 3D reconstruction

1. Introduction

Electron microscopy (EM) allows imaging of large biological macromolecules nearly in their native state. In single-particle reconstruction, EM projection images are taken from a specimen that shows neither a highly structured spatial distribution on the micrograph (as in 2D-crystal reconstruction, see Ruprecht and Nield, 2001), nor a marked internal symmetry (as in icosahedral particle reconstruction, see Mancini et al., 1997). In order to avoid radiation damage, the specimen is imaged at very low electron doses, resulting in extremely noisy projection images (signal-to-noise ratios are in the order of 0.1). This requires the combination of large numbers (10^3–10^5) of projection images in the process of volume reconstruction, typically resulting in 3D structures with resolutions between 6 and 30 Å. This range of resolution allows discerning biologically relevant information regarding molecular shape, domain architecture, and, on the high-resolution end, secondary structure (Frank, 2002).

The typical image processing steps needed to obtain a 3D reconstruction of a macromolecule may be briefly summarized as follows. After digitization of the electron micrographs, individual particle projections are identified. Then, images are classified to distinguish possible

* Corresponding author. Fax: +34-91-585-4506.
E-mail address: coss@cnb.uam.es (C.O.S. Sorzano).

1047-8477/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2004.06.006

structural variability, different projection directions, or simply, particles other than the one under study. Once a homogeneous projection set is achieved, the relative projection orientations (angular assignment) are determined. Subsequently, the 2D projections are combined into a 3D structure that is compatible with all the projections. High-resolution studies typically require the use of a high number of projections and the correction for image deterioration caused by the microscope image formation characteristics.

Success of EM single-particle analysis has been highly correlated with methodological advances and the continuous development of various EM image processing packages. Several comprehensive software packages for single-particle reconstruction are available to the electron microscopist, e.g., Spider (Frank et al., 1996), Imagic (van Heel et al., 1996), or Eman (Ludtke et al., 1999). Eight years ago, we introduced X-windows-based microscopy image processing package (Xmipp) to the EM community (Marabini et al., 1996) (http://www.cnb.uam.es/~bioinfo). Since then, a large amount of methodological advances has been added to Xmipp. Furthermore, we have completely re-written Xmipp in C++, now representing a hierarchical structure of well-documented data structures and functions. This modular design offers a convenient environment for testing and implementation of novel methodological developments. Only in the last year, Xmipp has been used in more than a dozen structural studies (Boskovic et al., 2003; Ferreira-Pereira et al., 2003; Gómez-Lorenzo et al., 2003; Hamada et al., 2003; Jouan et al., 2003; Llorca et al., 2003; Messaoudi et al., 2003; Oliva et al., 2003; Peng et al., 2003; Rizzo et al., 2003; Scheuring et al., 2003; Schleiff et al., 2003; Zickermann et al., 2003).

Xmipp is oriented to the complete processing of EM single particles, from image acquisition to 3D reconstruction. It provides a large set of individual programs to cover the different steps in the image processing work-flow. Fig. 1 provides a representative work-flow of the analysis tools covered by Xmipp. In order to ensure maximum compatibility with other available software packages, Xmipp native file formats are those of the commonly used Spider package, except for the selection files which can be converted easily. Due to the modular design of Xmipp, the user can choose to use either Xmipp or any other package at almost any point in the processing sequence. In the following we focus on those aspects of the package that clearly distinguish Xmipp from other packages in the field. These aspects are:

- *2D and 3D image classification*, with a special effort on neural-network based methods. In this context we have developed self-organizing maps (SOMs) with well-understood and controlled mathematical properties (Section 3.1).
- *Contrast transfer function (CTF)* estimation and incorporation into the reconstruction algorithm. We have developed tools to estimate the CTF using a 2D parametric model for the power spectral density followed by a 2D adjustment of a theoretical CTF model (Section 3.2). Once the CTF is estimated, Xmipp provides several methods for its correction (Section 3.3).
- *Angular assignment.* We have developed a robust angular assignment algorithm based on a multiresolution correlation with a reference volume. We have also proposed to refine the angular estimates using an accurate model of the parameter space (Section 3.4).
- *3D reconstruction methods.* We have explored iterative algorithms in which the basis functions used to describe the volumes are not voxels but smooth functions called *blobs* (see Section 3.5.1). Furthermore, constraints such as symmetry, total protein mass, macromolecular surface or positivity can be incorporated within the reconstruction process (Section 3.5.2). These constraints will be referred to as *volumetric constraints*.
- *Crystal reconstruction in real space*. We have developed a new algorithm for 3D reconstruction of 2D crystals, whose main departure from the traditional method is that it works in real space. This has the advantage of making it convenient to introduce volumetric constraints. (Section 3.5.3).
- *Parallel implementation* of the most time consuming algorithms (Section 3.5.4). In particular, we have shown that the reconstruction algorithms can be speeded up nearly linearly with the number of processors without compromising their convergence properties.
- *Objective comparison of 3D reconstruction methods*. The analysis of the relative performance of two different algorithms is a key task in image processing. Xmipp offers several approaches to assign a statistical significance level to the claim of superiority of one method over the other (Section 3.5.5).

The paper is organized as follows: Section 2 provides a general overview of the architecture and philosophy of the package. In Section 3 we discuss the new methodological developments implemented in Xmipp.

2. Xmipp architecture

Xmipp was originally meant for single-particle reconstruction. It has been designed to be fully compatible with the well-known Spider package (Frank et al., 1996). Not only the file formats for images, volumes, and document files are the same, but special care has been paid to conventions as axes orientations, Euler

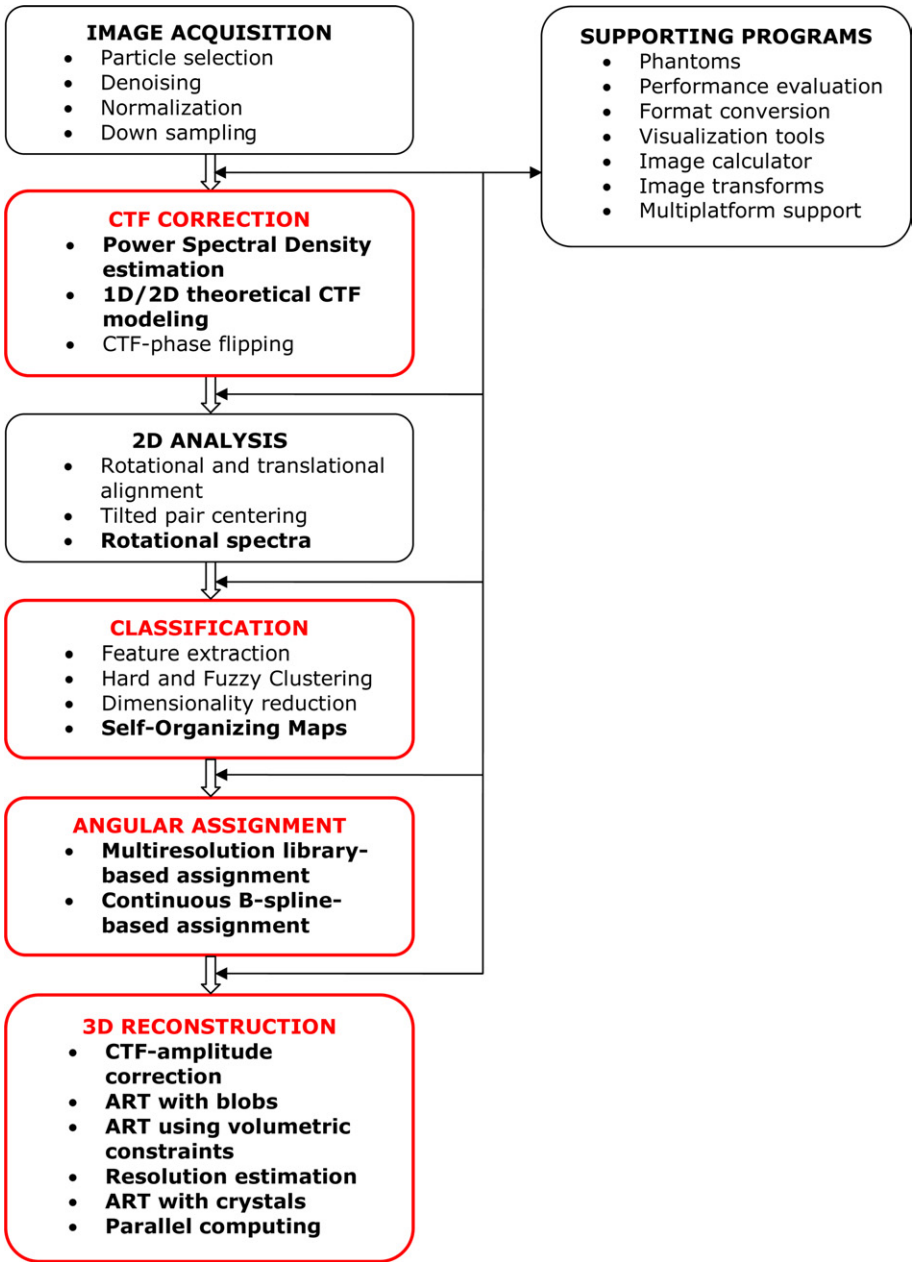

Fig. 1. Schematic work-flow of Xmipp image processing analysis. Those areas at which Xmipp introduces original contributions are depicted in red and bold font. It is noticeable that at any point in the analysis work-flow, the user can obtain or supply its own set of images. This feature makes Xmipp very flexible for its use in a combined analysis using other software packages. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

angles, image and volume centers, etc. (see "Conventions" and "File Formats" on the Programmer's site of Xmipp home page) In fact, positions 50–76 in the Spider header have been kindly reserved for our special needs. Compatibility with other software packages (e.g., EMAN, IMAGIC, MRC, etc.) can be obtained using image converters as *em2em* (http://www.imagescience.de/em2em) or *bimg* (http://www.niams.nih.gov/rcn/labbranch/lsbr/software/bsoft/bimg.html). Therefore, and due to the modular design of Xmipp, at almost any step in the Xmipp work flow the user can choose to enter or exit from/to the program package of his personal preference.

Xmipp is portable to any machine with the GNU C++ compiler and Qt graphical libraries. In practice, this includes virtually any UNIX-like platform and Windows via Cygwin. In particular, we routinely run it on IRIX, Linux, AIX, Solaris, and OSF machines. A full hierarchy of C++ classes implements all the package functionality (see "Libraries" on the Programmers' site of Xmipp home page). These classes are organized into separate libraries with a highly modular design. This

allows component re-use and quick prototyping of new image processing algorithms, making new developments easier to implement. Fig. 2 (top) shows a schematic overview of the Xmipp hierarchy.

The core of Xmipp is formed by function libraries: a data structure library, a classification library, and a reconstruction library, among others. The data structure library provides general data structures and functionalities such as vectors, matrices, volumes, Fourier transforms, Xmipp I/O files (Spider format, selection files, and document files), error functions, time functions, random functions, filename handling, type conversions, and command line reading. The classification library handles data structures and functionalities to support the implemented classification algorithms. Training and code vectors are among the data structures available in this library. Its functionalities include hard and fuzzy clustering, partitional clustering, principal component analysis, and SOMs. The reconstruction library implements all functionalities related to obtaining a 3D structure, including functions for angular assignment, contrast transfer function detection and correction, 3D reconstruction, and reconstruction-quality evaluation. This hierarchy of data structures and functions provides developers with a strong foundation for creating and testing new ideas related to Image Processing. The libraries are richly documented at the Programmer's site of Xmipp home page.

The function libraries serve the more than 100 separate programs currently contained in Xmipp. These programs are merely front-ends to the core libraries and form the interface to the user through the UNIX command-line, where all necessary input can be provided. This provides a large range of tools for inter-process communication and automation. Xmipp has been developed taking into account final users and, therefore, an effort has been made to present a user-friendly interface. Advanced users

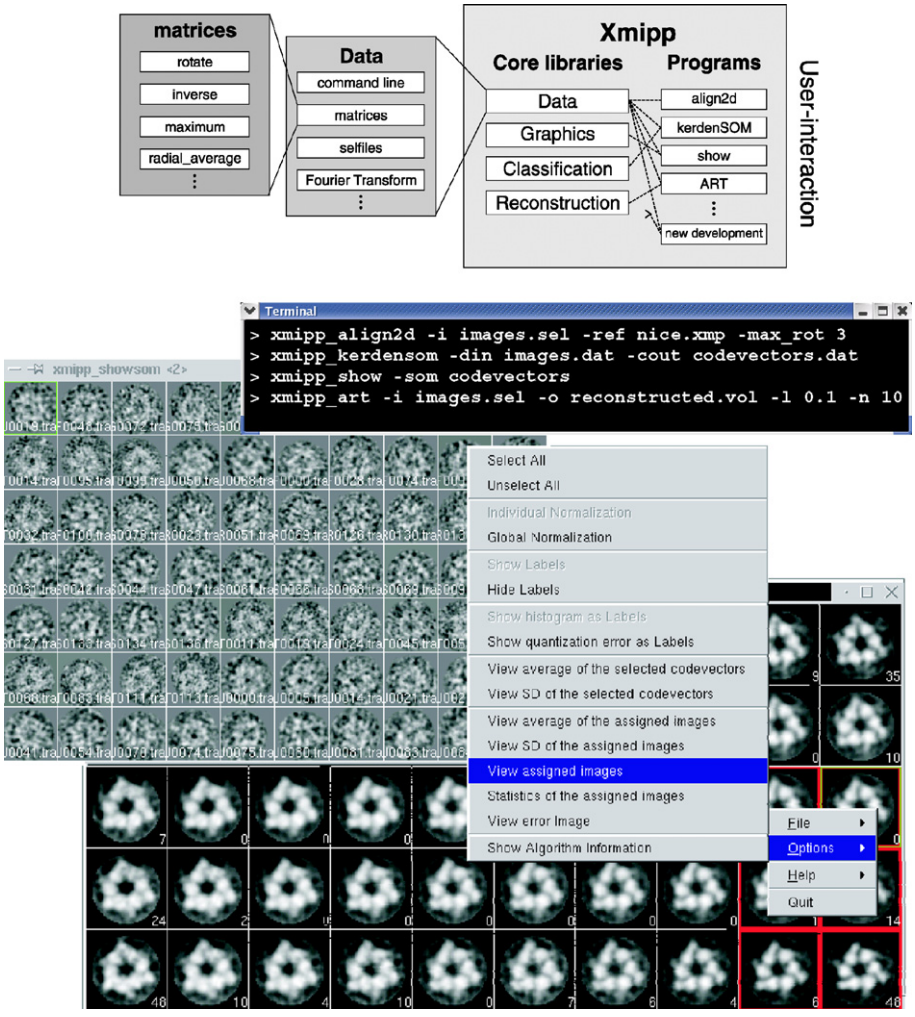

Fig. 2. (Top) The Xmipp function hierarchy encapsulates low-level objects into libraries, and these form the core library. User programs are front-ends to the package libraries. This programming strategy facilitates code re-usability and development of new algorithms. (Bottom) Example of a Graphical User Interface for the visualization and manipulation of SOMs and other clustering algorithms, and a command-line interface for the user programs.

may improve productivity by taking advantage of the scripting possibilities of the UNIX environments. Besides command-line user-interaction, graphical tools are available for particle picking and visualization of images, volumes, and SOMs (see Fig. 2 (bottom)).

Since Xmipp is a public-domain package, free access to the source code is granted. The most up-to-date version can be downloaded from the well-established public-domain repository SourceForge (http://www.sourceforge.net), that stores more than 74000 projects at this moment. The package comes with an extended documentation in the form of HTML pages. The same documentation is available on-line on the Xmipp home page (http://www.cnb.uam.es/~bioinfo). The main two sections are the ''User's site'' and the ''Programmer's site.'' The former contains manual pages for all user-end programs, including examples of their most common use. This part of the documentation has been written to serve the experimentalist while processing his data, or when getting acquainted with the Xmipp package. The ''Programmer's site'' provides detailed descriptions of all library functions and data structures, as well as information about conventions and file formats. This part of the documentation has been written to support developers in implementing new functionalities in Xmipp. In addition, the Xmipp documentation includes a comprehensive set of tutorial pages, each explaining different stages of the image processing trajectory. By means of elaborate examples and guided demos, these pages provide an extensive introduction, both to electron microscopy image processing in general, as to the use of the Xmipp-package in particular.

3. New methodological developments

3.1. Classification algorithms

Image classification is vital as a preprocessing step in EM. The goal is to sort the original population of images into different homogeneous sub-populations. This classification process produces a double benefit: (i) it helps to distinguish projection data coming from different specimens or from different realizations of the same specimen; (ii) it groups projections with similar projection directions.

3.1.1. Self-organizing maps

Xmipp approaches for classification are mainly based on the use of a special type of neural network known as SOMs (Kohonen, 1982) which possesses very interesting properties for data mapping and classification. This type of neural network was first introduced in the EM field by Marabini and Carazo (1994). Especially in cases with large numbers of noisy and high-dimensional data, like EM projection images, SOMs are excellent tools for classification. A SOM can be interpreted as a non-linear projection of high-dimensional data onto a regular low-dimensional grid [for a comprehensive review, see Kohonen (1997, 1998)]. Thereby, the large input data set is reduced to a relatively small set of representatives with highly improved signal-to-noise ratios (called code vectors). These representatives have the property of being ordered on a 2D grid tending to preserve the variance and topological characteristics of the input data. Subsequently, classification can be achieved by visual inspection of the output set of representatives.

Regarding Xmipp, the most important recent development in this area has been the introduction of the KerDenSOM (Kernel Density Estimator Self-Organizing Map) algorithm. This is a new SOM method based on a cost-function explicitly designed to obtain a set of code vectors whose probability density resembles the probability density of the input data (Pascual-Montano et al., 2001). The advantages over classical SOM rely on its precise mathematical formulation, producing an algorithm able to achieve both self-organization and probability estimation in very precise terms. In classification of single-particle projection images, the main challenge is to find structural heterogeneities hidden in the data. KerDenSOM has proved its effectiveness in elucidating the complex clustering structure of sets of noise EM projection images, where other classical methods failed (Gómez-Lorenzo et al., 2002; Pascual-Montano et al., 2001, 2002).

3.1.2. Other mapping and clustering algorithms

In addition to classical SOM as proposed by Kohonen, Xmipp also provides a set of new variants of this method that possess several interesting characteristics deserving more attention:

- *BatchSOM:* this algorithm implements a batch training of the classical Kohonen SOM. In the traditional sequential training, samples are presented to the map one at a time, and the algorithm gradually moves the representative vectors towards them. In the batch training, the data set is presented to the SOM as a whole, and the new representative items are weighted averages of the data vectors. Both algorithms are iterative, but the batch version has a faster convergence.
- *FuzzySOM:* this algorithm implements a different SOM, based on a new cost function and its optimization algorithm (Pascual-Marqui et al., 2001). The cost function is derived by introducing two modifications to Bezdek's generalized fuzzy c-means functional, which is widely used in cluster analysis and pattern recognition (Bezdek, 1981). In this way, a fuzzy SOM is obtained.

Xmipp also includes a more conventional battery of pattern recognition and statistical methods. Several programs implement some well-known linear and non-line-

ar mapping methods that are useful for reducing the dimensionality of the data set. These include linear mapping methods like principal component analysis (PCA) (van Heel and Frank, 1981) and non-linear techniques like Sammon mapping (Sammon, 1969). These mapping methods allow the projection of high-dimensional data into a lower dimensionality space in such a way that they can be explored and visualized to detect the clustering structure of the data set.

In addition, classical clustering methods already applied in the EM field are also available in Xmipp: Fuzzy c-Means (Carazo et al., 1990), Kernel c-Means (de Alarcón et al., 2002; Pascual-Montano et al., 2001), and Fuzzy Kohonen Clustering Network (Pascual et al., 2000).

Classification algorithms can be applied not only to single-particle images, but to any type of data in general. For example, a widely used feature extraction algorithm used in EM is the rotational spectrum (Crowther and Amos, 1971). Classification of rotational spectra allows the detection of rotational symmetry heterogeneities in a more accurate way than using pixels directly. A set of programs to calculate the rotational spectra from a set of single-particle images is also provided in this package.

Another example is the possibility to use the classification programs with volumes. This is particularly important in the case of classification of 3D electron tomograms where the first successful applications have already been reported (Pascual-Montano et al., 2002).

3.1.3. Exploratory data analysis

The ordered nature of the code vectors in SOM justifies its use as a display for data items: nearby code vectors will have similar data mapped onto them, so the density of code vectors will be an approximate reflection of the density of the input data. Therefore, the visual inspection of code vectors helps in understanding the intrinsic nature of the variability of the data set. A set of visualization programs in Xmipp (see Fig. 2 (bottom)) provides a functionality aiming at these visual inspection task needed in most EM structural studies, specially in those cases where no prior knowledge about the structure and variability of the data set is available.

3.2. CTF detection and phase correction

The effect of microscope aberrations in the images is described (in Fourier space) by the so-called contrast transfer function. The CTF filters both the high and the low frequencies, introduces zones of alternate contrast and eliminates all information at certain frequencies (Thon, 1966; Toyoshima et al., 1993; Unwin, 1973; Wade, 1992). Therefore, it is vital to correct for its effects if meaningful reconstructions are to be obtained.

Before correcting for the CTF, the CTF itself has to be estimated from the electron micrographs. This is usually done by adjusting a theoretical CTF model to the power spectral density (PSD) of the micrograph (Gao et al., 2002; Huang et al., 2003; Mindell and Grigorieff, 2003; Saad et al., 2001; Zhou et al., 1996).

A fully automatic method of assigning a theoretical CTF to each projection image is implemented in Xmipp. Instead of estimating the PSD using classical periodogram averaging, Xmipp employs a parametric PSD-estimation, using the so-called auto-regressive moving-average (ARMA) models. These models assume that the value of each pixel can be approximated by the values of its neighboring ones, plus a correlated noise term. In our experiments, this method resulted in clear Thon rings for the typically difficult PSD-estimation of non-carbon coated holey grid cryo-micrographs, while periodogram averaging yielded only faint and noisy rings (Velázquez-Muriel et al., 2003). Usually, the poor results of periodogram averaging are enhanced in other packages by radial symmetrization of the PSD estimation. However, this operation may hide slightly astigmatic situations and cannot be applied if strong astigmatism is present. An additional advantage of ARMA models compared to periodogram averaging is that typically fewer samples are needed to obtain a reliable PSD-estimation. This allows a more local estimation, which may be especially important when large CTF-variations are observed along the micrograph (e.g., for tilted ones).

In Xmipp the parameters of the well-known theoretical CTF-model (Zhou et al., 1996) is adjusted to the PSD estimated using ARMA modeling. Besides the classical 1D models, Xmipp also provides support for 2D ones. In this way, non-astigmatic as well as astigmatic images can be handled properly.

The output of this stage is a 2D estimation of the CTF for each particle projection. At this point, an easy CTF correction that has shown to be very powerful (Frank, 1996; Sorzano et al., 2004b) is the phase flipping. Flipping the CTF phase at an early stage is advantageous since it facilitates subsequent image processing. Xmipp provides programs to perform phase flipping on the experimental images.

3.3. CTF amplitude correction

If high-resolution reconstructions are to be achieved, the amplitude modulation introduced by the CTF must be corrected (Gabashvili et al., 2000; Ludtke et al., 2001; Mancini et al., 2000; Mueller et al., 2000). Xmipp uses the iterative data refinement (IDR) algorithm for this correction (Sorzano et al., 2004b). IDR modifies iteratively the experimental projections so that successive iterations are less and less affected by the amplitude modulation. The new projections are used for 3D reconstruction. This iterative process has been shown to

converge in practical cases to a CTF-free reconstruction.

Another possibility to correct for the CTF is to explicitly incorporate the corresponding point spread function (PSF) into the reconstruction equations. The problem of this approach is that the matrix of the equation system associated to the tomographic problem becomes not so sparse and the speed of the program is severely affected. As an alternative, we explored the Chahine algorithm (Zubelli et al., 2003). The underlying idea is not to solve the system of equations with the PSF explicitly taken into account but its normalized equation system, avoiding in this way the loss of efficiency. In fact, its convergence should be faster than that of IDR. Currently, Xmipp provides both methods for CTF amplitude correction, and a careful comparison between the two algorithms is underway.

3.4. Angular assignment

A key knowledge for performing a 3D reconstruction from 2D projections is the pose of each image, i.e., the point of view from which each one of the projections is taken. A common approach consists on comparing the experimental projections with the theoretical projections of a reference volume (reference library of projections) known to be similar to the macromolecule under study (Penczek et al., 1994; Radermacher, 1994). However, due to the high level of noise in the images it is likely that some of the particles are misassigned (Sorzano et al., 2004a). Even using true reference volumes, the proportion of misassigned particles can be as high as 98% in certain cases (Sorzano et al., 2004a).

Xmipp includes an angular assignment algorithm that deviates from other methods used in the field, mainly due to its multiresolution approach and its angular selection strategy. In this approach, correlation coefficients between experimental images and the library of all reference projections are computed in a coarse-to-fine fashion using a discrete wavelet transform. The experimental image is first compared to the reference library at a coarse resolution. Those reference images showing a high correlation with the experimental image progress to further comparison stages where the resolution is increased. This process is iterated until the full image resolution is achieved. Besides reducing computational costs, the gradual increase of information provides robustness with respect to noise. Additionally, the assignment of the angles is not based on merely selecting the correlation maximum (as in other approaches), but sets of several, highly correlated reference images are analyzed for each experimental image. This algorithm has been shown to be more robust than the standard angular assignment algorithms (Sorzano et al., 2004a).

The accuracy of any library-based algorithm is compromised by the computational cost of increasing the number of reference projections. Xmipp provides an answer to this problem by optimizing an objective function involving the angular and translational parameters of the experimental images (Jonić et al., 2003). The similarity between these images and the reference is optimized in a continuous way in Fourier-space, utilizing B-splines to model both the experimental projections and the reference volume (Unser et al., 1991). Angles assigned by the multiresolution algorithm (or any other method) may be used to initialize the optimization process. The combination of a library-based angular assignment method with a continuous refinement allows performing complete angular searches in very short times. At the same time, the robustness of the multiresolution angular-assignment algorithm is enhanced by the accurate assignment provided by the continuous refinement.

3.5. Three-dimensional reconstruction

3.5.1. ART + blobs

Xmipp employs *series expansion methods* for volume reconstruction. In these methods, the volume to be reconstructed is represented as a weighted sum of basis functions, reducing the reconstruction problem to the estimation of their weights. This is done in an iterative fashion so that the projections of the reconstructed volume computed through an image formation model resemble the experimental projections obtained by the microscope.

Xmipp implements different variants of iterative 3D reconstruction algorithms: algebraic reconstruction technique (ART, Herman, 1980), simultaneous iterative reconstruction technique (SIRT, Gilbert, 1972), simultaneous algebraic reconstruction technique (SART, Andersen and Kak, 1984), component averaging (CAV, Censor et al., 2001a), block-iterative component averaging (BICAV, Censor et al., 2001b), and Averaging Strings (Censor et al., 2000). For the choice of the basis functions, we followed the approach proposed by Lewitt (1990, 1992), using Kaiser–Bessel window functions (usually referred to as blobs) instead of voxels. Blobs have the property of being smooth in real and Fourier space, making them especially suitable for working under the very noisy conditions of EM projection images. The choice of the grid (the set of localizations of the blob centers) has also been shown to influence the quality of the reconstruction. It has been found that grids different from the simple cubic grid (SC grid) are preferable for placing these blobs. In particular the body centered cubic grid (BCC grid) proves to be the best one (Garduño and Herman, 2001; Matej and Lewitt, 1995). Xmipp allows to reconstruct in a SC, BCC or face centered cubic (FCC) grid. We reach the conclusion that the so-called algebraic reconstruction technique (ART)

is the optimal choice for many cases in the EM single-particle reconstruction field (Fernández et al., 2002; Marabini et al., 1998; Marabini et al., 1997; Sorzano et al., 2001).

The validity and strength of this approach has been illustrated in multiple experimental 3D studies that have employed ART + blobs for volume reconstruction, (e.g., Area et al., 2004; Bárcena et al., 2001; Gómez-Lorenzo et al., 2003; Llorca et al., 1999; San Martín et al., 1998).

3.5.2. Volumetric constraints

A major advantage of real-space reconstruction techniques like ART, is that they provide a relatively simple way of imposing real-space constraints on the reconstructed volumes. There exist many situations where information about the volume is known a priori. This type of information may include non-negativity, total mass, knowledge about molecular symmetry, and also knowledge of the molecular surface. The latter may be obtained using atomic-force or metal-shadowing microscopy (Engel and Muller, 2000; Fuchs et al., 1995), or non-linear image processing (Sorzano et al., 2002b). All these types of information provide restrictions on the reconstructed volume, which can be expressed as volumetric constraints. Such constraints can be formulated as a new equation system to be solved at the same time as the one imposed by the experimental projections (Sorzano et al., 2002b). The Xmipp implementation of ART allows such application of constraints to the reconstructed volumes. Our experiments show that these constraints are particularly useful in cases where projection data is missing in certain directions (e.g., missing cone), since the volumetric constraints help to fill these empty regions in the projection space (Sorzano et al., 2002a,b).

3.5.3. Crystal reconstruction

Recently, the ART algorithm has been further modified to allow 3D-reconstruction of 2D crystals in real-space (Marabini et al., 2002, 2004). The main motivation for this work was not to replace the existing, well-established methods for 2D crystal reconstruction (Crowther et al., 1996), but to provide a framework in which prior information about the surface relief, as mentioned above, could be applied in the 2D-crystal case (Dimmeler et al., 2001). However, our tests indicate that the real-space algorithm behaves better than the traditional one, even without the incorporation of such additional information, especially when the number of projections is small. As the number of projections increases, the results of the two algorithms become more similar (Marabini et al., 2004).

3.5.4. Parallel computing in 3D reconstruction

EM single-particle analysis is characterized by its huge computational demands, both in computing time and memory requirements. Many EM laboratories therefore have clusters of workstations as part of their facilities, and use parallel computing for the most time-consuming stages of structure determination (Fernández et al., 2002; Marinescu et al., 2001; Peltier et al., 2003). The current implementation of Xmipp includes parallel support for 3D reconstruction by iterative real-space techniques (SIRT, SART, CAV, BICAV, and Averaging Strings) of both single-particles and single-tilt electron tomography data. The implemented parallel strategies make use of the message-passing interface (MPI, Gropp et al., 1999), a standard in parallel-programming, and are based on *domain decomposition* and the SPMD model (*Single Program, Multiple Data*) in which every node in the system carries out, essentially, the same task over its own data domain.

For the electron tomography case, speedups nearly linear with the number of nodes have been obtained (Fernández et al., 2002). However, although parallel strategies allow great reductions of computing time per iteration, they may negatively affect the convergence rate of the reconstruction algorithm. Therefore, special care must be taken in order to avoid losing overall performance (Bilbao-Castro et al., 2003).

3.5.5. Working with phantoms

The analysis of the relative performance of two different algorithms is a key task in image processing. Such an analysis is typically non-trivial, and a serious comparison should not rely on visual inspection of a few images. Xmipp provides programs for helping the developers in this task. These programs design phantom volumes (either from geometrical elements such as cylinders, blobs, cones, etc. or from PDB atomic structures), generate projections with any data collection geometry, add several kinds of noise, simulate the CTF effect, etc. In order to assess the quality of the different algorithms, their output can be compared with the original phantom (that is, the ideal reconstruction) via numerical observers called figures of merit (FOMs). Xmipp is particularly rich concerning the number and variety of available FOMs.

The use of phantoms and FOMs for the fine tuning of algorithms has proved to be a powerful tool for performing structural studies (Sorzano et al., 2001). However, the analysis of the huge amount of FOMs under varying experimental conditions constitutes a challenging problem which we have addressed using multivariate statistical analysis (Sorzano et al., 2001).

4. Conclusions

In this paper we present an overview of the image processing software package Xmipp. This free, open-source software package, which is portable to most UNIX-like platforms, has recently undergone large

changes. A wide range of new methodologies for EM single-particle reconstruction has been implemented, and the package has been extended to include functionalities for electron tomography and 2D crystals. The novel methods that have been described deviate from the conventional techniques used in the field, which illustrates the surplus value of Xmipp for the EM community. In addition, its current implementation in a hierarchy of C++ classes provides excellent tools for the development and testing of new methodologies. This software structure facilitates Xmipp integration into future standardization efforts for EM software packages. Finally, we propose this new generation of Xmipp as an open invitation for both experimentalists and developers within the EM community.

Acknowledgments

We acknowledge partial support from the "Comunidad Autónoma de Madrid" through Grant CAM-07B-0032-2002, the "Comisión Interministerial de Ciencia y Tecnología" of Spain through Grants BIO2001-1237, BIO2001-4253-E, BIO2001-4339-E, and BIO2002-10855-E, the European Union through Grants QLK2-2000-00634, QLRI-2000-31237, QLRT-2000-0136, and QLRI-2001-00015, and the NIH through Grant HL70472. Partial support from the "Universidad San Pablo-CEU" Grant 17/02 is also acknowledged. We are thankful to Dr. Carmen San Martín from the Biocomputing Unit of the National Center for Biotechnology and Dr. J.J. Fernández from the Univ. de Almería for revising the manuscript.

References

Andersen, A., Kak, A., 1984. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. Ultrasonic Imag. 6, 81–94.

Area, E., Martín-Benito, J., Gastaminza, P., Torreira, E., Valpuesta, J., Carrascosa, J., Ortín, J., 2004. 3D structure of the influenza virus polymerase complex: localization of subunit domains. Proc. Natl. Acad. Sci. USA 101, 308–313.

Bárcena, M., Donate, L., Ruiz, T., Dixon, N., Radermacher, M., Carazo, J.M., 2001. The DnaB–DnaC complex: a structure based on interactions among assymetric dimers. EMBO J. 20, 1462–1468.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York.

Bilbao-Castro, J.R., Carazo, J.M., García, I., Fernández, J.J., 2003. Parallel iterative reconstruction methods for structure determination of biological specimens by electron microscopy. In: Proc. of IEEE Intl. Congress on Image Processing. vol. 1. pp. 565–568.

Boskovic, J., Rivera-Calzada, A., Maman, J., Chacón, P., Willison, K., Pearl, L., Llorca, O., 2003. Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. EMBO J. 22, 5875–5882.

Carazo, J.M., Rivera, F.F., Zapata, E.L., Radermacher, M., Frank, J., 1990. Fuzzy sets-based classification of electron microscopy images of biological macromolecules with an application to ribosomal particles. J. Microsc. 157, 187–203.

Censor, Y., Elfving, T., Herman, G., 2000. Averaging strings of sequential iterations for convex feasibility problems. In: Butnariu, D., Censor, Y., Reich, S. (Eds.), Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications. Elsevier Science, Amsterdam, pp. 101–114.

Censor, Y., Gordon, D., Gordon, R., 2001a. BICAV: A block-iterative parallel algorithm for sparse systems with pixel-related weighting. IEEE Trans. Med. Imag. 20, 1050–1060.

Censor, Y., Gordon, D., Gordon, R., 2001b. Component averaging: an efficient iterative parallel algorithm for large and sparse unstructured problems. Parallel computing 27, 777–808.

Crowther, R.A., Amos, L.A., 1971. Harmonic analysis of electron microscope images with rotational symmetry. J. Mol. Biol. 60, 123–130.

Crowther, R.A., Henderson, R., Smith, J.M., 1996. MRC image processing programs. J. Struct. Biol. 116, 9–16.

de Alarcón, P.A., Pascual-Montano, A., Gupta, A., Carazo, J.M., 2002. Modeling shape and topology of low-resolution density maps of biological macromolecules. Biophys. J. 83, 619–632.

Dimmeler, E., Marabini, R., Tittmann, P., Gross, H., 2001. Correlation of topographic surface and volume data from 3D electron microscopy. J. Struct. Biol. 136, 20–29.

Engel, A., Muller, D.J., 2000. Observing single biomolecules at work with the atomic force microscope. Nat. Struct. Biol. 7, 715–718.

Fernández, J., Lawrence, A., Roca, J., García, I, Ellisman, M., Carazo, J., 2002. High performance electron tomography of complex biological specimens. J. Struct. Biol. 138, 6–20.

Ferreira-Pereira, A., Marco, S., Decottignies, A., Nader, J., Goffeau, A., Rigaud, J., 2003. Three-dimensional reconstruction of the *Saccharomyces cerevisiae* multidrug resistance protein Pdr5p. J. Biological Chemistry 278, 11995–11999.

Frank, J., 1996. Three Dimensional Electron Microscopy of Macromolecular Assemblies. Academic Press, San Diego, CA.

Frank, J., 2002. Single-particle imaging of macromolecules by cryo-electron microscopy. Ann. Rev. Biophy. Biomol. Struct., 303–319.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., Leith, A., 1996. SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. 116, 190–199.

Fuchs, K.G., Tittmann, P., Krusche, K., Gross, H., 1995. Reconstruction and representation of surface data from two-dimensional crystalline biological macromolecules. Bioimaging 3, 12–24.

Gabashvili, I.S., Agrawal, R.K., Spahn, C.M., Grassucci, R.A., Svergun, D.I., Frank, J., Penczek, P., 2000. Solution structure of the *E. coli* 70S ribosome at 11.5Å resolution. Cell 100, 537–549.

Gao, H., Spahn, C., Grassucci, R., Frank, J., 2002. An assay for local quality in cryo-electro micrographs of single particles. Ultramicroscopy 93, 169–178.

Garduño, E., Herman, G.T., 2001. Optimization of basis functions for both reconstruction and visualization. In: Sébastien Fourey, G.T.H., Kong, T.Y. (Eds.), Electronic Notes in Theoretical Computer Science, vol. 46. Elsevier.

Gilbert, P., 1972. Iterative methods for the three-dimensional reconstruction of an object from projections. J. Theor. Biol. 36, 105–117.

Gómez-Lorenzo, M.G., Valle, M., Donate, L.E., Gruss, C., Bárcena, M., Sorzano, C.O.S., Frank, J., Carazo, J.M., 2002. Structural studies of SV40 large T-antigen on the origin of replication. In: Proc. of the 15th International Conference on Electron Microscopy (ICEM). Durban, South Africa.

Gropp, W., Lusk, E., Thakur, R., 1999. Using MPI: Portable Parallel Programming with the Message Passing Interface, second ed. MIT Press, Cambridge.

Gómez-Lorenzo, M., Valle, M., Frank, J., Gruss, C., Sorzano, C.O.S., Chen, X.S., Donate, L.E., Carazo, J.M., 2003. Large T antigen on

the simian virus 40 origin of replication: a 3D snapshot prior to DNA replication. EMBO J. 22, 6205–6213.
Hamada, K., Terauchi, A., Mikoshiba, K., 2003. Three-dimensional rearrangements within inositol 1,4,5-trisphosphate receptor by calcium. J. Biol. Chem. 278, 52881–52889.
Herman, G.T., 1980. Image Reconstruction from Projections: The Fundamentals of Computerized Tomography. Academic Press, New York.
Huang, Z., Baldwin, P.R., Mullapudi, S., Penczek, P.A., 2003. Automated determination of parameters describing power spectra of micrograph images in electron microscopy. J. Struct. Biol. 144, 79–94.
Jonić, S., Thévenaz, P., Unser, M., 2003. Multiresolution-based registration of a volume to a set of its projections. In: Proc. of the SPIE Symposium on Medical Imaging. vol. 5032. San Diego, pp. 1049–1052.
Jouan, L., Marco, S., Taveau, J., 2003. Revisiting the structure of *Alvinella pompejana* hemoglobin at 20 Å resolution by cryoelectron microscopy. J. Struct. Biol. 143, 33–44.
Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybernet. 43, 59–69.
Kohonen, T., 1997. Self-organizing Maps. Springer-Verlag, Berlin.
Kohonen, T., 1998. The self-organizing maps. Neurocomputing 21, 1–6.
Lewitt, R.M., 1990. Multidimensional digital image representations using generalized Kaiser-Bessel window functions. J. Opt. Soc. Am. A 7, 1834–1846.
Lewitt, R.M., 1992. Alternatives to voxels for image representation in iterative reconstruction algorithms. Phys. Med. Biol. 37, 705–716.
Llorca, O., Rivera-Calzada, A., Grantham, J., Willison, K., 2003. Electron microscopy and 3D reconstructions reveal that human ATM kinase uses an arm-like domain to clamp around double-stranded DNA. Oncogene 22, 3867–3874.
Llorca, O., Smyth, M.G., Carrascosa, J.L., Willison, K.R., Radermacher, R., Steinbacher, S., Valupuesta, J.M., 1999. 3D reconstruction of the ATP-bound form of CCT reveals assymetric folding conformation of a type II chaperonin. Nat. Struct. Biol. 7, 639–642.
Ludtke, S.J., Baldwin, P.R., Chiu, W., 1999. EMAN: semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. 128, 82–97.
Ludtke, S.J., Jakana, J., Song, J., Chuang, D.T., Chiu, W., 2001. A 11.5 single particle reconstruction of GroEL using EMAN. J. Mol. Biol. 314, 253–262.
Mancini, E., Clarke, M., Gowen, B., Rutten, T., Fuller, S., 2000. Cryo-electron microscopy reveals the functional organization of an enveloped virus, Semliki Forest virus. Mol. Cell 5, 255–266.
Mancini, E.J., de Haas, F., Fuller, S.D., 1997. High-resolution icosahedral reconstruction: fulfilling the promise of cryo-electron microscopy. Structure 5, 741–750.
Marabini, R., Carazo, J.M., 1994. Pattern recognition and classification of images of biological macromolecules using artificial neural networks. Biophys. J. 66, 1804–1814.
Marabini, R., Herman, G.T., Carazo, J.M., 1998. 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs). Ultramicroscopy 72, 53–65.
Marabini, R., Masegosa, I.M., San Martín, M.C., Marco, S., Fernández, J.J., de la Fraga, L.G., Vaquerizo, C., Carazo, J.M., 1996. Xmipp: An image processing package for electron microscopy. J. Struct. Biol. 116, 237–240.
Marabini, R., Rietzel, E., Schröder, R., Herman, G.T., Carazo, J.M., 1997. Three-dimensional reconstruction from reduced sets of very noisy images acquired following a single-axis tilt schema: application of a new three-dimensional reconstruction algorithm and objective comparison with weighted backprojection. J. Struct. Biol. 120, 363–371.
Marabini, R., Sorzano, C.O.S., Fernández, J.J., Matej, S., Carazo, J.M., Herman, G.T., 2002. 3D reconstruction of 2D crystals from projections in real space. In: Proc. of the 1st International Symposium on Biomedical Imaging. Washington, pp. 689–692.
Marabini, R., Sorzano, C.O.S., Matej, S., Fernández, J.J., Carazo, J.M., Herman, G.T., 2004. 3D reconstruction of 2D crystals. IEEE Trans. Image Process. 13, 549–561.
Marinescu, D., Ji, Y., Lynch, R., 2001. Space-time tradeoffs for parallel 3D reconstruction algorithms for virus-structure determination. Concurr. Comput.: Practice Experience 13, 1083–1106.
Matej, S., Lewitt, R.M., 1995. Efficient 3D grids for image reconstruction using spherically symmetric volume elements. IEEE Trans. Nuclear Sci. 42, 1361–1370.
Messaoudi, C., Boudier, T., Lechaire, J., Rigaud, J., Delacroix, H., Gaill, F., Marco, S., 2003. Use of cryo-negative staining in tomographic reconstruction of biological objects: application to T4 bacteriophage. Biol. Cell 95, 393–398.
Mindell, J.A., Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. J. Struct. Biol. 142, 334–347.
Mueller, R., Sommer, I., Baranov, P., Matadeen, R., Stoldt, M., Wohnert, J., Gorlach, M., van Heel, M., Brimacombe, R., 2000. The 3D arrangement of the 23S and 5S rRNA in the *Escherichia coli* 50S ribosomal subunit based on a cryo-electron microscopic reconstruction at 7.5 Å resolution. J. Molecular Biology 298, 35–59.
Oliva, M., Huecas, S., Palacios, J., Martin-Benito, J., Valpuesta, J., Andreu, J., 2003. Assembly of archaeal cell division protein FtsZ and a GTPase-inactive mutant into double-stranded filaments. J. Biol. Chem. 278, 33562–33570.
Pascual, A., Bárcena, M., Merelo, J.J., Carazo, J.M., 2000. Mapping and fuzzy classification of macromolecular images using self organizing neural networks. Ultramicroscopy 84, 85–99.
Pascual-Marqui, R.D., Pascual-Montano, A., Kochi, K., Carazo, J.M., 2001. Smoothly distributed fuzzy c-means: a new self-organizing map. Pattern Recognit. 34, 2395–2402.
Pascual-Montano, A., Donate, L.E., Valle, M., Bárcena, M., Pascual-Marqui, R.D., Carazo, J.M., 2001. A novel neural network technique for analysis and classification of EM single-particle images. J. Struct. Biol. 133, 233–245.
Pascual-Montano, A., Taylor, K.A., Winkler, H., Pascual-Marqui, R.D., Carazo, J.M., 2002. Quantitative self-organizing maps for clustering electron tomograms. J. Struct. Biol. 138, 114–122.
Peltier, S., Lin, A., Lee, D., Mock, S., Lamont, S., Molina, T., Wong, M., Dai, L., Martone, M., Ellisman, M., 2003. The telescience portal for tomography applications. J. Par. Distr. Comp. 63, 539–550.
Penczek, P.A., Grasucci, R.A., Frank, J., 1994. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. Ultramicroscopy 53, 251–270.
Peng, G., Fritzsch, G., Zickermann, V., Schagger, H., Mentele, R., Lottspeich, F., Bostina, M., Ra-dermacher, M., Huber, R., Stetter, K., Michel, H., 2003. Isolation, characterization and electron microscopic single particle analysis of the NADH:ubiquinone oxidoreductase (complex I) from the hyperthermophilic eubacterium *Aquifex aeolicus*. Biochemistry 42, 3032–3039.
Radermacher, M., 1994. Three-dimensional reconstruction from random projections-orientational alignment via radon transforms. Ultramicroscopy 53, 121–136.
Rizzo, V., Coskun, U., Radermacher, M., Ruiz, T., Armbruster, A., Gruber, G., 2003. Resolution of the VI ATPase from *Manduca sexta* into subcomplexes and visualization of an ATPase-active A3B3EG complex by electron microscopy. J. Biol. Chem. 278, 270–275.
Ruprecht, J., Nield, J., 2001. Determining the structure of biological macromolecules by transmission electron microscopy, single par-

ticle analysis and 3D reconstruction. Progr. Biophys. Mol. Biol. 75, 121–164.

Saad, A., Ludtke, S., Jakana, J., Rixon, F., Tsuruta, H., Chiu, W., 2001. Fourier amplitude decay of electron cryomicroscopic images of single particles and effects on structure determination. J. Struct. Biol. 133, 32–42.

Sammon, J.W., 1969. A nonlinear mapping for data structure analysis. IEEE Trans. Comput. 18, 401–409.

San Martín, M.C., Radermacher, M., Wolpensinger, B., Engel, A., Miles, C.S., Dixon, N.E., Carazo, J.M., 1998. Three-dimensional reconstructions from cryoelectron microscopy images reveal an intimate complex between helicase DnaB and its loading partner DnaC. Struct. Fold. Design 6, 501–509.

Scheuring, S., Seguin, J., Marco, S., Levy, D., Robert, B., Rigaud, J., 2003. Nanodissection and high-resolution imaging of the *Rhodopseudomonas viridis* photosynthetic core complex in native membranes by AFM. Atomic force microscopy. Proc. Natl. Acad. Sci. USA 100, 1690–1693.

Schleiff, E., Soll, J., Kuchler, M., Kuhlbrandt, W., Harrer, R., 2003. Characterization of the translocon of the outer envelope of chloroplasts. J. Cell Biol. 160, 541–551.

Sorzano, C.O.S., Jonic, S., El-Bez, C., Carazo, J.M., De Carlo, S., Thévenaz, P., Unser, M., 2004a. A multiresolution approach to pose assignment in 3-D electron microscopy of single particles. J. Struct. Biol. 146, 381–392.

Sorzano, C.O.S., Marabini, R., Boisset, N., Rietzel, E., Schröder, R., Herman, G.T., Carazo, J.M., 2001. The effect of overabundant projection directions on 3D reconstruction algorithms. J. Struct. Biol. 133, 108–118.

Sorzano, C.O.S., Marabini, R., Fernández, J.J., Velázquez-Muriel, J.A., Herman, G.T., Carazo, J.M., 2002a. New reconstruction conditions greatly improve the reconstruction quality. In: Proc. of the 15th International Conference on Electron Microscopy (ICEM). Durban, South Africa, pp. 431–432.

Sorzano, C.O.S., Marabini, R., Herman, G.T., Carazo, J.M., 2002b. Volumetric constraints in 3D tomography applied to electron microscopy. In: Proc. of the 1st International Symposium on Biomedical Imaging. Washington, USA, pp. 641–644.

Sorzano, C.O.S., Marabini, R., Herman, G.T., Censor, Y., Carazo, J.M., 2004b. Transfer function restoration in 3D electron microscopy via iterative data refinement. Phys. Med. Biol. 49, 509–522.

Thon, F., 1966. Zur defokussierungsabhängigkeit des phasenkontrastes bei der elektronen-mikroskopischen abbildung. Z. Naturforsch 21a, 476–478.

Toyoshima, C., Yonekura, K., Sasabe, H., 1993. Contrast transfer for frozen-hydrated specimens II: amplitude contrast at very low frequencies. Ultramicroscopy 48, 165–176.

Unser, M., Aldroubi, A., Eden, M., 1991. Fast B-spline transforms for continuous image representation and interpolation. IEEE Trans. Pattern Anal. Mach. Intell. 13, 277–285.

Unwin, P., 1973. Phase contrast electron microscopy of biological materials. J. Microsc. 98, 299–312.

van Heel, M., Frank, J., 1981. Use of multivariate statistical statistics in analysing the images of biological macromolecules. Ultramicroscopy 6, 187–194.

van Heel, M., Harauz, G., Orlova, E.V., Schmidt, R., Schatz, M., 1996. A new generation of the IMAGIC image processing system. J. Struct. Biol. 116, 17–24.

Velázquez-Muriel, J.A., Sorzano, C.O.S., Fernández, J.J., Carazo, J.M., 2003. A method for estimating the CTF in electron microscopy based on ARMA models and parameter adjusting. Ultramicroscopy 96, 17–35.

Wade, R., 1992. A brief look at imaging and contrast transfer. Ultramicroscopy 46, 145–156.

Zhou, Z.H., Hardt, S., Wang, B., Sherman, M.B., Jakana, J., Chiu, W., 1996. CTF determination of images of ice-embedded single particles using a graphics interface. J. Struct. Biol. 116, 216–222.

Zickermann, V., Bostina, M., Hunte, C., Ruiz, T., Radermacher, M., Brandt, U., 2003. Functional implications from an unexpected position of the 49-kDa subunit of NADH:ubiquinone oxidoreductase. J. Biol. Chem. 278, 29072–29078.

Zubelli, J.P., Marabini, R., Sorzano, C.O.S., Herman, G.T., 2003. Three-dimensional reconstruction by Chahine's method from electron microscopic projections corrupted by instrumental aberrations. Inverse Probl. 19, 933–949.